



# Quasi-linear SPDEs in divergence form

Felix Otto<sup>1</sup> · Hendrik Weber<sup>2</sup>

Received: 5 July 2017  
© The Author(s) 2018

## Abstract

We develop a solution theory in Hölder spaces for a quasi-linear stochastic PDE driven by an additive noise. The key ingredients are two deterministic PDE lemmas which establish a priori Hölder bounds for a parabolic equation in divergence form with irregular right-hand-side term. We apply these bounds to the case of a right-hand-side noise term which is white in time and trace class in space, to obtain stretched exponential bounds for the Hölder semi-norms of the solution.

**Keywords** Quasi-linear SPDEs · Regularity theory · Hölder spaces

## 1 Introduction

We are interested in the quasi-linear equation

$$\partial_t u - \nabla \cdot A(\nabla u) = \xi, \quad (1)$$

with unknown  $u: \mathbb{R}_t \times \mathbb{R}_x^d \rightarrow \mathbb{R}$  for a nonlinearity  $A: \mathbb{R}^d \rightarrow \mathbb{R}^d$  that is uniformly elliptic. The right hand side  $\xi$  represents an irregular distribution; the key example we have in mind is a noise term which is “white in time” and “coloured in space”. The aim of this article is to develop a priori bounds in Hölder spaces leading to a solution theory for (1).

The regularity of the noise terms appearing in stochastic differential equations is often effectively measured on the Hölder scale. This is well known in the finite-dimensional case, the most classical example being Brownian motion, which has (locally)  $\alpha$ -Hölder continuous trajectories for any  $\alpha < \frac{1}{2}$ . Statements in other scales of spaces, e.g. in  $L^2$ -based fractional Sobolev spaces are possible but are weaker: Brownian trajectories almost surely take values in  $H_{loc}^\alpha$  for  $\alpha < \frac{1}{2}$ , but this does not

---

✉ Hendrik Weber  
hendrik.weber@warwick.ac.uk

<sup>1</sup> Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, 04103 Leipzig, Germany

<sup>2</sup> University of Warwick, Coventry, UK

even imply the continuity of trajectories. It thus seems natural to seek a solution theory in Hölder spaces also for stochastic partial differential equations.

In the case of semi-linear equations such a theory is by now classical and well-developed, see e.g. [1,3,11]. For example, in the case of the stochastic heat equation

$$\partial_t v - \Delta v = \xi, \quad (2)$$

the variation-of-constants formula leads to an explicit representation of  $v$  in terms of the heat kernel (the so-called “mild solutions”) which can be used to deduce optimal Hölder bounds. This approach extends to equations with lower-order non-linearities such as stochastic reaction–diffusion equations or the stochastic Navier–Stokes equation.

In the case of the quasi-linear equations we consider, there is no natural mild formulation of the equation. However, equations such as (1) have been treated since the 70’s (see e.g. the classical works [5,9] or [10] for a more recent presentation) using a “variational formulation”, which relies on the theory of monotone operators and yields solutions that satisfy

$$\sup_{0 \leq t \leq T} \int u^2(t, x) dx + \int_0^T \int |\nabla u(t, x)|^2 dx dt < \infty \quad (3)$$

for all  $T < \infty$  almost surely. In fact, these methods allow for much more general equations; generalisations include degenerate cases such as the porous medium equation.

The aim of the present article is to demonstrate how purely deterministic PDE arguments can be used to improve on the energy inequality (3) and obtain estimates on space–time Hölder norms of  $\nabla u$ . Our main deterministic result, Corollary 1, states, roughly speaking, that we can bound the (parabolic) Hölder semi-norm  $[\nabla u]_\alpha$  for solutions of  $u$  of (1) in terms of the corresponding semi-norm  $[\nabla v]_\alpha$  for solutions of the linear problem (2). The proof splits into Lemma 1 where this bound is established for a small  $\alpha_0$  using the celebrated De Giorgi–Nash Theorem, and into Lemma 2 where it is upgraded to arbitrary  $\alpha$  by Schauder theory. The techniques employed follow classical PDE arguments, as developed for example in [6,7], but they have to be adjusted to the low-regularity right hand side.

To illustrate the implications of our deterministic result in the case of random  $\xi$ , we treat the case where  $\xi$  is a Gaussian distribution that is white in time and coloured in space. This type of noise is commonly studied in the literature, often using the “differential” notation

$$du = \nabla \cdot A(\nabla u) dt + dW, \quad (4)$$

where  $W$  is a Wiener process with spatial covariance operator  $K$ . Our assumption on  $\xi$  corresponds to saying that  $K$  is a trace-class operator, which is precisely the assumption needed in the variational approach. We restrict ourselves to the case where  $\xi$  is periodic and compactly supported in time. This assumption is made to yield bounds on  $\nabla v$  which hold uniformly over space and time. The only stochastic ingredient of

this article is Lemma 3, where Gaussian moments for  $[\nabla v]_\alpha$  are established, using the covariance of  $\xi$  and its Gaussianity. Theorem 1 combines the main deterministic result, Corollary 1, and Lemma 3 to construct spatially periodic solutions  $u$  with zero initial data, i.e.  $u|_{t \leq 0} = 0$ . We establish existence and uniqueness of solutions to (1) in Theorem 1, as well as stretched exponential moments for  $[\nabla u]_\alpha$ .

Our result is closely related to the recent work [2], where a Hölder theory for the quasi-linear stochastic PDE

$$du = \nabla \cdot (A(u)\nabla u)dt + H(u)dW \quad (5)$$

is developed. The first step of that work is to consider the auxiliary equation

$$dv = \Delta v dt + H(u)dW.$$

The authors use some a priori information in the spirit of the energy estimate (3) as well as martingale inequalities to get a priori control on  $\nabla v$ . The key observation in their approach is that this a priori control on  $v$  allows to rewrite the equation for the remainder  $w = u - v$  as

$$\partial_t w = \nabla \cdot (A(u)\nabla w) + \nabla \cdot ((A(u) - I)\nabla v)$$

and to obtain Hölder regularity for  $w$  using the De Giorgi–Nash Theorem. We pursue a similar strategy, and work with the equation for  $w = u - v$ . However, (1) is more non-linear than (5) and the classical PDE results presented in [6, 7] do not immediately apply in this low-regularity situation. Our main deterministic result, Corollary 1, provides the necessary bound.

In a previous version of this work, see [8], we treated a quasi-linear equation

$$\frac{1}{T}\bar{u} + \partial_t \bar{u} - \partial_x^2 \pi(\bar{u}) = \bar{\xi}, \quad (6)$$

where  $\bar{\xi}$  is a space–time white noise over  $\mathbb{R}_t \times \mathbb{R}_x$ , and derived a stretched exponential moment bound akin to (24) on the Hölder semi-norms  $[u]_\alpha$ . The results in the present article contain this result, up to the different treatment of large scales. Indeed, specialising (1) to the case  $d = 1$  and differentiating with respect to  $x$  yields for  $\bar{u} = \partial_x u$

$$\partial_t \bar{u} - \partial_x^2 A(\bar{u}) = \partial_x \xi,$$

which coincides with (6), noting that our assumptions on  $\xi$  cover the case where  $\bar{\xi} = \partial_x \xi$  is a space–time white noise in one spatial dimension, and that in the one-dimensional case our assumptions on  $A$  coincide with the assumptions imposed on  $\pi$  in [8]. The key difference between the approach proposed in [8] and the approach we present here is that the core arguments are now purely deterministic and the use of log-Sobolev inequalities can be fully avoided.

## 2 Setting

For the deterministic part of our paper we rewrite the noise term  $\xi$  as  $\partial_t v - \Delta v$  where  $v$  solves (2). We thus strive to get bounds of solutions to

$$\partial_t u - \nabla \cdot A(\nabla u) = \partial_t v - \Delta v, \quad (7)$$

in terms of  $v$ . Here and throughout the paper we interpret equations in the distributional sense over all of  $\mathbb{R}_t \times \mathbb{R}_x^d$ . In order to stress the divergence form of the right hand side, we relabel those terms and write

$$\partial_t u - \nabla \cdot A(\nabla u) = \partial_t v + \nabla \cdot j, \quad (8)$$

for  $j = -\nabla v$ . We present a Schauder theory where we estimate the solution  $\nabla u$  by the data  $(v, j)$  in the Hölder space  $C^\alpha$ , always with respect to the parabolic distance

$$d((t', x'), (t, x)) := \sqrt{|t' - t|} + |x' - x|. \quad (9)$$

This is slightly different from the standard Schauder theory in  $C^{1,\alpha}$ , which cannot be applied due to the right-hand-side term  $\partial_t v$  that is irregular in time. In fact we shall control the  $C^{1,\alpha}$ -semi norm of  $w := u - v$

$$[w]_{1+\alpha} := [\nabla w]_\alpha + \sup_{t,t',x} \frac{|w(t, x) - w(t', x)|}{\sqrt{|t - t'|}^{\alpha+1}}, \quad (10)$$

where  $[\cdot]_\alpha$  denotes the (parabolic) Hölder semi-norm on space-time  $\mathbb{R}_t \times \mathbb{R}_x^d$

$$[w]_\alpha := \sup_{z' \neq z \in \mathbb{R} \times \mathbb{R}^d} \frac{|w(z') - w(z)|}{d^\alpha(z', z)}. \quad (11)$$

We make two assumptions on the nonlinearity  $A: \mathbb{R}^d \rightarrow \mathbb{R}^d$  in form of assumptions on the tensor field given by the derivative matrix  $DA$ :

**Assumption 1**  $DA$  is uniformly elliptic in the sense that there exists a constant  $\lambda > 0$  such that

$$\xi \cdot DA(q)\xi \geq \lambda |\xi|^2 \quad \text{and} \quad |DA(q)\xi| \leq |\xi| \quad \text{for all vectors } q, \xi. \quad (12)$$

Here, without loss of generality we normalized the upper bound to unity.

We will make use of (12) in the following form: For every spatial shift vector  $y \in \mathbb{R}^d$  we will work with the increment operator  $\delta_y u(t, x) = u(t, x + y) - u(t, x)$  and use the chain-type rule

$$\delta_y A(\nabla u) = a_y \delta_y \nabla u \quad (13)$$

where

$$a_y(t, x) = \int_0^1 DA(\theta \nabla u(t, x + y) + (1 - \theta) \nabla u(t, x)) d\theta.$$

Then (12) ensures that for all  $y$  we have uniform ellipticity of  $a_y$ :

$$\eta \cdot a_y(t, x) \eta \geq \lambda |\eta|^2 \quad \text{and} \quad |a_y(t, x) \eta| \leq |\eta| \quad \text{for all } (t, x), \eta. \quad (14)$$

**Assumption 2**  $DA$  is globally Lipschitz in the sense that there exists a constant  $\Lambda < \infty$  such that

$$|DA(q') - DA(q)| \leq \Lambda |q' - q| \quad \text{for all } q, q'. \quad (15)$$

We will make use of (15) in the following form: For any exponent  $\beta \in (0, 1]$  we have the following estimate on the level of Hölder norms

$$[a_y]_\beta \leq \Lambda [\nabla u]_\beta \quad \text{and} \quad [A(\nabla u)]_\beta \leq \Lambda [\nabla u]_\beta. \quad (16)$$

We use Eq. (8) exclusively in the following form: We apply the increment operator  $\delta_y$  to it and obtain by (13)

$$\partial_t \delta_y u - \nabla \cdot a_y \nabla \delta_y u = \partial_t \delta_y v + \nabla \cdot \delta_y j,$$

which in terms of the difference  $w := u - v$  we rewrite as

$$\partial_t \delta_y w - \nabla \cdot a_y \nabla \delta_y w = \nabla \cdot (a_y \nabla \delta_y v + \delta_y j). \quad (17)$$

We establish our form of  $C^{1,\alpha}$ -Schauder theory, cf. Corollary 1, in two lemmas. While the Lemma 1 just relies on the uniform ellipticity (12) and crucially uses the  $C^\alpha$ -a priori estimate for  $\delta_y w$  of De Giorgi and Nash based on (17), Lemma 2 uses also the Lipschitz continuity (15) and proceeds by a more standard Schauder-type argument.

**Lemma 1** *There exists an exponent  $\alpha_1 = \alpha_1(d, \lambda) \in (0, 1)$  such that for any exponent  $\alpha_0 \in (0, \alpha_1)$  we have*

$$[\nabla u]_{\alpha_0} \leq C(d, \lambda, \alpha_0) ([\nabla v]_{\alpha_0} + [j]_{\alpha_0}), \quad (18)$$

*provided we already have the qualitative information that the left hand side is finite.*

The critical point in the proof of Lemma 1 is that we extract control of  $[\nabla w]_{\alpha_0}$  (and thus  $[\nabla u]_{\alpha_0}$ ) from (17) without having to pass to the limit in the difference (quotient)  $\delta_y$ , which is not possible due to the low regularity of  $\nabla v$ .

**Lemma 2** *Let  $\alpha_0$  be as in Lemma 1 and suppose that  $L$  is so small that*

$$[\nabla u]_{\alpha_0, P_{2L}} \leq L^{-\alpha_0}, \quad (19)$$

where  $P_R := (-R^2, 0) \times B_R$  denotes the (centered) parabolic cylinder of size  $R$  and  $[\cdot]_{\beta, P_R}$  the  $\beta$ -Hölder semi-norm restricted to this set. Then we have for any exponent  $\alpha \in [\alpha_0, 1)$

$$[\nabla u]_{\alpha, P_L} \leq C(d, \lambda, \Lambda, \alpha_0, \alpha) (L^{-\alpha} + [\nabla v]_{\alpha, P_{2L}} + [j]_{\alpha, P_{2L}}). \quad (20)$$

**Corollary 1** *Let  $\alpha_0$  be as in Lemma 1. Then we have for any exponent  $\alpha \in (0, 1)$*

$$\begin{aligned} [u - v]_{1+\alpha} + [\nabla u]_{\alpha} \\ \leq C(d, \lambda, \Lambda, \alpha) \left( ([\nabla v]_{\alpha_0} + [j]_{\alpha_0})^{\frac{\alpha}{\alpha_0}} + ([\nabla v]_{\alpha} + [j]_{\alpha}) \right), \end{aligned}$$

provided we already have the qualitative information that  $[\nabla u]_{\alpha_0} < \infty$ .

To illustrate an application of Corollary 1, we treat the case where the right hand side is a stochastic noise which is white in time but coloured in space. Such a noise term is described by a Gaussian random distribution  $\xi$  over  $(t, x) \in \mathbb{R} \times \mathbb{R}^d$ , the probability distribution of which is characterized by having zero mean and

$$\langle (\xi, \varphi)^2 \rangle = \int_0^1 \int \int K(x, y) \varphi(t, x) \varphi(t, y) dx dy dt,$$

where  $(\xi, \varphi)$  stands for  $\xi$  tested against the Schwartz function  $\varphi \in \mathcal{S}(\mathbb{R} \times \mathbb{R}^d)$  and  $\langle \cdot \rangle$  is used for the expectation of a random variable. The spatial correlation  $K$  can be seen as the kernel of a regularising operator. Such a noise term is standard in the SPDE literature, often written in “differential notation” as

$$(\xi, \varphi) = \int \langle \varphi(t, \cdot), dW_t \rangle,$$

where  $W$  is an  $L^2$ -valued Wiener process with covariance operator  $K$ , see e.g. [1, Sect. 5].

Denote by  $v$  the solution of the constant-coefficient heat Eq. (2). Under suitable conditions on the kernel  $K$  it is known that  $\nabla v$  is regular enough, i.e.  $\alpha$ -Hölder continuous, to apply the above deterministic theory. As illustration we treat the case where  $\xi$  is assumed to be 1-periodic in all spatial directions, say of period 1, and in addition localised to a compact time interval, say the interval  $[0, 1]$ . If we assume in addition that the probability distribution of  $\xi$  is translation invariant in the spatial directions, so that  $K(x, y) = K(x - y)$ , we have the following convenient Fourier series representation

$$\xi(t, x) = \sum_{k \in (2\pi\mathbb{Z})^d} e^{ik \cdot x} \sqrt{\hat{K}(k)} \mathbf{1}_{[0, 1]}(t) \dot{\beta}_k(t). \quad (21)$$

Here the  $\beta_k$  are complex-valued standard Brownian motions (i.e. real and imaginary parts are independent and satisfy  $\langle \Re(\beta_k(t))^2 \rangle = \langle \Im(\beta_k(t))^2 \rangle = \frac{t}{\sqrt{2}}$ ), that are independent up to the constraint  $\beta_k = \bar{\beta}_{-k}$ , which ensures that  $\xi$  is real-valued, and  $\dot{\beta}_k(t)$  stands for the distributional time derivative. The  $\hat{K}(k)$  are real-valued, non-negative and symmetric in the sense that  $\hat{K}(k) = \hat{K}(-k)$ . The almost sure convergence of (21) in the space of distributions can be easily shown, but we adopt the slightly simpler framework to only work with  $v$ , which we define by its Fourier series representation:

$$v(t, x) := \sum_{k \in (2\pi\mathbb{Z})^d} \sqrt{\hat{K}(k)} e^{ik \cdot x} \int_0^{\min\{t, 1\}} e^{-(t-s)|k|^2} d\beta_k(s). \quad (22)$$

In order to ensure that the gradient is well behaved we impose that there exists  $s > d$  such that for  $k \in (2\pi\mathbb{Z})^d$

$$\hat{K}(k) \leq (1 + |k|^2)^{-\frac{s}{2}}, \quad (23)$$

where we have set the normalisation equal to 1 without loss of generality. Incidentally, this condition on  $s$  precisely says that the spatial covariance operator  $K$  is of trace class. Then we have the following lemma.

**Lemma 3** *Let  $v(t, x)$  be given by (22) for  $t > 0$  and set  $v_{t \leq 0} = 0$ . Then for  $\alpha < \min\{\frac{s-d}{2}, 1\}$  there exists  $C_0 = C_0(d, \alpha, s) < \infty$  such that*

$$\left\langle \exp \left( \frac{1}{C_0} [\nabla v]_\alpha^2 \right) \right\rangle < \infty,$$

where  $\langle \cdot \rangle$  denotes the expectation with respect to the probability distribution of  $v$ .

Combining our main deterministic result, Corollary 1, with the stochastic result in Lemma 3 we arrive at the following theorem.

**Theorem 1** *Let  $A$  be uniformly elliptic with ellipticity contrast  $\lambda$  and let  $DA$  be Lipschitz continuous with constant  $\Lambda$  (in the sense of (12) and (15)). Let  $\alpha_0 = \alpha_0(d, \lambda)$  be as in Lemma 1. Let  $v$  be given by (22) for a covariance operator  $K$  satisfying (23) for some  $s > d$ . Then for almost all realisations of  $v$ , there exists a unique  $u = u(t, x)$  with the following properties:*

- $u$  is continuous, 1-periodic in all spatial directions (i.e.  $u(t, x) = u(t, x + k)$  for all  $k \in \mathbb{Z}^d$ ) and  $u|_{t \leq 0} = 0$ .
- $[\nabla u]_\alpha, [u - v]_{1+\alpha} < \infty$  for  $\alpha < \min\{\frac{s-d}{2}, 1\}$ .
- $u$  solves (7) in the distributional sense, i.e. for all Schwartz functions  $\varphi \in \mathcal{S}(\mathbb{R} \times \mathbb{R}^d)$

$$\begin{aligned} & - \int \int \partial_t \varphi u dx dt + \int \int \nabla \varphi \cdot A(\nabla u) dx dt \\ & = - \int \int \partial_t \varphi v dx dt + \int \int \nabla \varphi \cdot \nabla v dx dt. \end{aligned}$$

Furthermore, for  $\alpha < \min\{\frac{s-d}{2}, 1\}$  there exists  $C = C(d, \lambda, \Lambda, \alpha, s) < \infty$  such that

$$\left\langle \exp \left( \frac{1}{C} ([\nabla u]_\alpha + [u - v]_{1+\alpha})_\alpha^{2 \min\{1, \frac{\alpha_0}{\alpha}\}} \right) \right\rangle < \infty, \quad (24)$$

where  $\langle \cdot \rangle$  denotes the expectation with respect to the probability distribution of  $v$ .

### 3 Proof of Theorem 1

Throughout this proof we use the symbol  $\lesssim$  for  $\leq C(d, \lambda, \Lambda, \alpha, s)$ . All functions  $u, v, w$  etc. appearing in the proof are assumed to be one-periodic in all space directions.

We assume we are given continuous functions  $v$  and  $j$  with  $[\nabla v]_\alpha, [j]_\alpha < \infty$  for an  $\alpha \in (0, 1)$ , which are 1-periodic in each spatial direction and with  $v|_{t \leq 0} = j|_{t \leq 0} = 0$ . We show that there exists a unique function  $u$  which is one-periodic in each spatial direction, satisfies  $u|_{t \leq 0} = 0$  and which satisfies

$$\begin{aligned} & - \int \int \partial_t \varphi u dx dt + \int \int \nabla \varphi \cdot A(\nabla u) dx dt \\ & = - \int \int \partial_t \varphi v dx dt - \int \int \nabla \varphi \cdot j dx dt, \end{aligned} \quad (25)$$

for each Schwartz function  $\varphi$ . In addition we show the bound

$$[\nabla u]_\alpha + [u - v]_{1+\alpha} \lesssim N, \quad (26)$$

where  $N = ([\nabla v]_\alpha + [j]_\alpha)^{\frac{\alpha}{\alpha_0}} + ([\nabla v]_\alpha + [j]_\alpha)$ . The desired existence and uniqueness statement then follows, by applying this to the case where  $v$  is given by (22),  $j = -\nabla v$ . For (24) we combine (26) and Lemma 3 to get for a suitable  $C = C(d, \lambda, \Lambda, \alpha, s)$

$$\begin{aligned} & \left\langle \exp \left( \frac{1}{C} ([\nabla u]_\alpha + [u - v]_{1+\alpha})_\alpha^{2 \min\{1, \frac{\alpha_0}{\alpha}\}} \right) \right\rangle \\ & \leq \left\langle \exp \left( \frac{1}{C_0} ([\nabla v]_\alpha)^{\frac{\alpha}{\alpha_0}} + [\nabla v]_\alpha^{2 \min\{1, \frac{\alpha_0}{\alpha}\}} \right) \right\rangle < \infty. \end{aligned}$$

The existence of solutions follows by approximation through regularisation. Let  $j_\varepsilon, v_\varepsilon$  be space–time regularisations (e.g. by convolution with suitable smooth kernel) of  $j, v$  satisfying  $[j_\varepsilon]_\alpha \leq [j]_\alpha, [\nabla v_\varepsilon]_\alpha \leq [\nabla v]_\alpha$  and such that  $v_\varepsilon|_{t \leq -\varepsilon} = j_\varepsilon|_{t \leq -\varepsilon} = 0$ . Then by classical theory there exists a unique classical solution  $u_\varepsilon$  for

$$\partial_t u_\varepsilon - \nabla \cdot A(\nabla u_\varepsilon) = \partial_t v_\varepsilon - \nabla \cdot j_\varepsilon, \quad u_\varepsilon|_{t \leq -\varepsilon} = 0,$$

which is one-periodic in all spatial directions (see e.g. [7, Thm. 12.14] for a proof in the case of Dirichlet data on a bounded spatial domain. The case of the torus is only simpler). In this situation Corollary 1 applies and yields



$$\begin{aligned} & [\nabla u_\varepsilon]_\alpha + [u_\varepsilon - v_\varepsilon]_{1+\alpha} \\ & \lesssim \left( ([\nabla v_\varepsilon]_\alpha + [j_\varepsilon]_\alpha)^{\frac{\alpha}{\alpha_0}} + ([\nabla v_\varepsilon]_\alpha + [j_\varepsilon]_\alpha) \right) \lesssim N. \end{aligned} \quad (27)$$

This estimate together with the initial datum  $u_\varepsilon|_{t=-\varepsilon} = v_\varepsilon|_{t=-\varepsilon} = 0$  permit to apply the Arzelà–Ascoli Theorem and to conclude that up to choosing a subsequence  $u_\varepsilon - v_\varepsilon \rightarrow w$ ,  $\nabla(u_\varepsilon - v_\varepsilon) \rightarrow \nabla w$ ,  $\nabla u_\varepsilon \rightarrow \nabla u$  locally uniformly for functions  $u, w$  with  $u = w + v$ . Furthermore,  $w$  solves

$$\partial_t w - \nabla \cdot A(\nabla u) = \nabla \cdot j$$

in the distributional sense. Setting  $u = w + v$  we obtain (25) and the estimate (26) follows by passing to the limit in (27) using lower semi-continuity.

It only remains to argue for (pathwise) uniqueness. Assume thus that  $u^1$  and  $u^2$  are one-periodic in space, satisfy (25) and vanish for  $t \leq 0$ . Thus the difference  $\delta u := u^1 - u^2$  satisfies

$$\partial_t \delta u = \nabla \cdot (A(\nabla u_1) - A(\nabla u_2)), \quad (28)$$

in the distributional sense and  $\delta u|_{t=0} = 0$ . In order to show that  $\delta u = 0$  we aim to test Eq. (28) against  $\delta u$  to obtain the identity

$$\frac{1}{2} \int_{[0,1]^d} \delta u^2(T, x) dx = - \int_0^T \int_{[0,1]^d} \nabla \delta u \cdot (A(\nabla u_1) - A(\nabla u_2)) dx dt, \quad (29)$$

for all  $T \geq 0$ . Once the identity (29) is justified, we can invoke the uniform ellipticity (14) once more and obtain the point-wise identity

$$\begin{aligned} \nabla \delta u \cdot (A(\nabla u_1) - A(\nabla u_2)) &= \nabla \delta u \cdot \left( \int_0^1 DA(\lambda \nabla u_1 + (1-\lambda) \nabla u_2) d\lambda \right) \nabla \delta u \\ &\geq 0 \end{aligned}$$

so that (29) yields  $\delta u = 0$ .

It thus remains to justify (29). For this we convolve (28) with a temporal regularising kernel at scale  $\varepsilon$  and then test against  $\delta u_\varepsilon$ , the temporally regularised version of  $\delta u$ . Here we use the fact that under the periodicity assumption the weak formulation (25) can be restated equivalently by replacing the space integrals over  $\mathbb{R}^d$  by integrals over  $[0, 1]^d$  and assuming that the test functions are also periodic. This yields for any  $T > 0$

$$\frac{1}{2} \int_{[0,1]^d} \delta u_\varepsilon^2(T, x) dx = - \int_{-\infty}^T \int_{[0,1]^d} \nabla \delta u_\varepsilon \cdot (A(\nabla u_1) - A(\nabla u_2))_\varepsilon dx dt. \quad (30)$$

We can pass to the limit  $\varepsilon \rightarrow 0$  on both sides using the fact that  $\delta u = (u^1 - v) - (u^2 - v)$  is  $\frac{1+\alpha}{2}$ -Hölder continuous in time and using the fact that  $\nabla u^1$  and  $\nabla u^2$  are  $\frac{\alpha}{2}$ -Hölder continuous in time.

## 4 Proof of Lemma 1

Throughout this proof we write  $\lesssim$  for  $\leq C(d, \lambda, \alpha_0)$ .

Based on (17) and (14) we have by a localized version of the Hölder a priori estimate of De Giorgi and Nash that there exists an exponent  $\alpha_1 = \alpha_1(d, \lambda) \in (0, 1)$  such that for all shift vectors  $y$ , all length scales  $\ell$  and all space–time points  $z$

$$[\delta_y w]_{\alpha_1, P_\ell(z)} \lesssim \ell^{-\alpha_1} \inf_k \|\delta_y w - k\|_{P_{2\ell}(z)} + \ell^{1-\alpha_1} \|a_y \nabla \delta_y v + \delta_y j\|_{P_{2\ell}(z)}, \quad (31)$$

where  $P_\ell(z) = (t - \ell^2, t) \times B_\ell(x)$  denotes the parabolic cylinder centered around  $z = (t, x)$ , and where  $\|\cdot\|_{P_\ell(z)}$  stands for the supremum norm restricted to the set  $P_\ell(z)$ . The exponents of the  $\ell$ -factors in (31) are determined by scaling; smuggling in the constant  $k$  is possible since (14) is oblivious to changing  $\delta_y w$  by an additive constant. We refer to [7, Theorem 6.28] as one possible reference (with  $b \equiv 0$ ,  $c^0 \equiv 0$ , and  $g \equiv 0$  so that  $k_1 = \sup_{Q(R)} |f|$  in the notation of that reference). We fix an exponent  $\alpha_0 \in (0, \alpha_1)$  and take the supremum of (31) over all shift vectors  $y$  with  $|y| \leq r$  for some  $r \leq \ell$

$$\begin{aligned} & \sup_{|y| \leq r} [\delta_y w]_{\alpha_1, P_\ell(z)} \\ & \lesssim \ell^{-\alpha_1} \sup_{|y| \leq r} \inf_k \|\delta_y w - k\|_{P_{2\ell}(z)} + \ell^{1-\alpha_1} \sup_{|y| \leq r} \|a_y \nabla \delta_y v + \delta_y j\|_{P_{2\ell}(z)}. \end{aligned} \quad (32)$$

We first estimate the right-hand-side terms of (32). We start with the second right-hand-side term: From the definition (11) of the Hölder semi-norm and that of the parabolic cylinder, we obtain

$$\begin{aligned} & \|a_y \nabla \delta_y v + \delta_y j\|_{P_{2\ell}(z)} \\ & \stackrel{(14)}{\leq} \|\delta_y \nabla v\|_{P_{2\ell}(z)} + \|\delta_y j\|_{P_{2\ell}(z)} \leq |y|^{\alpha_0} ([\nabla v]_{\alpha_0} + [j]_{\alpha_0}), \end{aligned}$$

so that

$$\ell^{1-\alpha_1} \sup_{|y| \leq r} \|a_y \nabla \delta_y v + \delta_y j\|_{P_{2\ell}(z)} \lesssim \ell^{1-\alpha_1} r^{\alpha_0} ([\nabla v]_{\alpha_0} + [j]_{\alpha_0}). \quad (33)$$

We now turn to the first right-hand-side term of (32): We first note that

$$\inf_k \|\delta_y w - k\|_{P_{2\ell}(z)} \leq r \inf_c \|\nabla w - c\|_{P_{3\ell}(z)}, \quad (34)$$

where the right-hand-side infimum ranges over all  $c \in \mathbb{R}^d$ . Indeed, passing to  $\tilde{w}(t, x) = w(t, x) - c \cdot y$ , so that  $\nabla w - c = \nabla \tilde{w}$ , and transforming  $\tilde{k} = k - c \cdot y$ , so that  $\delta_y w - k = \delta_y \tilde{w} - \tilde{k}$ , we see that (34) reduces to  $\|\delta_y \tilde{w}\|_{P_{2\ell}(z)} \leq |y| \|\nabla \tilde{w}\|_{P_{3\ell}(z)}$ , which because of  $|y| \leq r \leq \ell$  is a consequence of the mean-value theorem. Since obviously  $\inf_c \|\nabla w - c\|_{P_{3\ell}(z)} \leq (3\ell)^{\alpha_0} [\nabla w]_{\alpha_0}$ , we obtain

$$\ell^{-\alpha_1} \sup_{|y| \leq r} \inf_k \|\delta_y w - k\|_{P_{2\ell}(z)} \lesssim \ell^{\alpha_0 - \alpha_1} r [\nabla w]_{\alpha_0}. \quad (35)$$

We finally turn to the left hand side term in (32) and note

$$\sup_{|y| \leq r} [\delta_y w]_{\alpha_1, P_\ell(z)} \geq \sup_{|y| \leq r} [\delta_y w]_{\alpha_1, P_r(z)} \geq \frac{1}{r^{\alpha_1}} \sup_{|y| \leq r} \inf_k \|\delta_y w - k\|_{P_r(z)}. \quad (36)$$

Inserting (33), (35), and (36), into (32) we obtain

$$\begin{aligned} & \frac{1}{r^{\alpha_1}} \sup_{|y| \leq r} \inf_k \|\delta_y w - k\|_{P_r(z)} \\ & \lesssim \ell^{\alpha_0 - \alpha_1} r [\nabla w]_{\alpha_0} + \ell^{1 - \alpha_1} r^{\alpha_0} ([\nabla v]_{\alpha_0} + [j]_{\alpha_0}), \end{aligned}$$

which we multiply with  $\frac{1}{r^{1 + \alpha_0 - \alpha_1}}$  to arrive at

$$\begin{aligned} & \frac{1}{r^{1 + \alpha_0}} \sup_{|y| \leq r} \inf_k \|\delta_y w - k\|_{P_r(z)} \\ & \lesssim \left(\frac{r}{\ell}\right)^{\alpha_1 - \alpha_0} [\nabla w]_{\alpha_0} + \left(\frac{\ell}{r}\right)^{1 - \alpha_1} ([\nabla v]_{\alpha_0} + [j]_{\alpha_0}). \end{aligned} \quad (37)$$

We now argue that we are done once we establish the norm equivalence

$$[\nabla w]_{\alpha_0} \lesssim \sup_{z, r} \frac{1}{r^{1 + \alpha_0}} \sup_{|y| \leq r} \inf_k \|\delta_y w - k\|_{P_r(z)}. \quad (38)$$

Indeed, choosing  $\ell = Mr$  with  $M \geq 1$  to be chosen later, we take the supremum of (37) over all radii  $r$  and all space–time points  $z$  to arrive at

$$\begin{aligned} & \sup_{z, r} \frac{1}{r^{1 + \alpha_0}} \sup_{|y| \leq r} \inf_k \|\delta_y w - k\|_{P_r(z)} \\ & \lesssim M^{\alpha_0 - \alpha_1} [\nabla w]_{\alpha_0} + M^{1 - \alpha_1} ([\nabla v]_{\alpha_0} + [j]_{\alpha_0}), \end{aligned}$$

into which we insert (38)

$$[\nabla w]_{\alpha_0} \lesssim M^{\alpha_0 - \alpha_1} [\nabla w]_{\alpha_0} + M^{1 - \alpha_1} ([\nabla v]_{\alpha_0} + [j]_{\alpha_0}).$$

By the triangle inequality in  $[\cdot]_{\alpha_0}$  we post-process this to

$$[\nabla u]_{\alpha_0} \lesssim M^{\alpha_0 - \alpha_1} [\nabla u]_{\alpha_0} + M^{1 - \alpha_1} ([\nabla v]_{\alpha_0} + [j]_{\alpha_0}).$$

Since by our qualitative assumption of  $[\nabla u]_{\alpha_0} < \infty$ , and since  $\alpha_0 < \alpha_1$ , we may choose  $M = M(d, \lambda, \alpha_0)$  so large that this turns into the desired (18).

We now turn to the norm equivalence (38); the elements of the argument are standard in modern Schauder theory, in the spirit of [4, Theorem 3.3.1]. By rotational symmetry, it is enough to establish

$$[\partial_1 w]_{\alpha_0} \lesssim \sup_{z,r} \frac{1}{r^{1+\alpha_0}} \sup_{|y| \leq r} \inf_k \|\delta_y w - k\|_{P_r(z)} =: N. \quad (39)$$

Let  $k = k(y, r, z)$  denote the optimal constant in the right hand side of (39). We first argue that for arbitrary but fixed point  $z$ , we have for all radii  $r$

$$|k(2re_1, 2r, z) - 2k(re_1, r, z)| \lesssim Nr^{1+\alpha_0}. \quad (40)$$

Indeed, based on the telescoping identity  $\delta_{2re_1} w = \delta_{re_1} w + \delta_{re_1} w(\cdot + re_1)$  we obtain by the triangle inequality the following additivity of  $k$  in the  $y$ -variable

$$\begin{aligned} |k(2re_1, 2r, z) - 2k(re_1, 2r, z)| &\leq \|\delta_{2re_1} w - k(2re_1, 2r, z)\|_{P_r(z)} \\ &\quad + \|\delta_{re_1} w - k(re_1, 2r, z)\|_{P_r(z)} + \|\delta_{re_1} w(\cdot + re_1) - k(re_1, 2r, z)\|_{P_r(z)} \\ &\leq \|\delta_{2re_1} w - k(2re_1, 2r, z)\|_{P_{2r}(z)} + 2\|\delta_{re_1} w - k(re_1, 2r, z)\|_{P_{2r}(z)} \\ &\leq 3(2r)^{1+\alpha_0} N. \end{aligned}$$

Likewise, we have that  $k$  only mildly depends on the  $r$ -variable

$$\begin{aligned} |k(re_1, 2r, z) - k(re_1, r, z)| &\leq \|\delta_{re_1} w - k(re_1, 2r, z)\|_{P_{2r}(z)} + \|\delta_{re_1} w - k(re_1, r, z)\|_{P_r(z)} \\ &\leq 2(2r)^{1+\alpha_0} N. \end{aligned}$$

From the two last estimates, we obtain (40). Since  $\alpha_0 > 0$ , we learn from (40) that there exists a constant  $c_1(z)$  such that

$$\left| \frac{1}{r} k(re_1, r, z) - c_1(z) \right| \lesssim Nr^{\alpha_0},$$

along a given dyadic sequence of radii  $r$ . We insert this into the definition of  $N$  to obtain

$$\left\| \frac{1}{r} \delta_{re_1} w - c_1(z) \right\|_{P_r(z)} \lesssim Nr^{\alpha_0},$$

from which, since in particular  $u$  and thus  $w$  is differentiable in the spatial variable, we learn that  $c_1(z) = \partial_1 w(z)$  so that

$$\left\| \frac{1}{r} \delta_{re_1} w - \partial_1 w(z) \right\|_{P_r(z)} \lesssim Nr^{\alpha_0}. \quad (41)$$

Since we identified the limit, this now holds for any radius  $r$  (and not just the dyadic ones). Given two points  $z, z'$  we set  $r := 2d(z, z')$ , cf. (9), and obtain

$$\begin{aligned} & |\partial_1 w(z) - \partial_1 w(z')| \\ & \leq \left\| \frac{1}{r} \delta_{re_1} w - \partial_1 w(z) \right\|_{P_{\frac{r}{2}}(z)} + \left\| \frac{1}{r} \delta_{re_1} w - \partial_1 w(z') \right\|_{P_{\frac{r}{2}}(z')} \\ & \leq \left\| \frac{1}{r} \delta_{re_1} w - \partial_1 w(z) \right\|_{P_r(z)} + \left\| \frac{1}{r} \delta_{re_1} w - \partial_1 w(z') \right\|_{P_r(z')}. \end{aligned}$$

Hence (39) follows from (41).

## 5 Proof of Lemma 2

Throughout this proof we use  $\lesssim$  for  $\leq C(d, \lambda, \Lambda, \alpha_0, \alpha)$ .

Let the two scales  $r \leq \ell \leq \frac{\ell}{4}$  be arbitrary and for the time being fixed. Let  $y$  be an arbitrary shift vector with  $|y| \leq r$ . By (16) in the localized form of  $[a_y]_{\alpha_0, P_{3\ell}}$   $\leq \Lambda[\nabla u]_{\alpha_0, P_{3\ell+r}}$  and (19) we have

$$[a_y]_{\alpha_0, P_{3\ell}} \lesssim \ell^{-\alpha_0}. \quad (42)$$

In conjunction with (14) we see that we may apply standard  $C^{1, \alpha_0}$ -Schauder theory to the parabolic operator  $\partial_t - \nabla \cdot a_y \nabla$  when localized to  $P_{3\ell}$ . We learn from rescaling according to  $(t, x) = (\ell^2 \hat{t}, \ell \hat{x})$  that (42) is exactly the control on the coefficient needed so that the constant in this localized Schauder theory is of the desired form  $C(d, \lambda, \Lambda, \alpha_0, \alpha)$ . We refer to [7, Theorem 4.8] for a possible reference (with  $b \equiv 0$ ,  $c \equiv 0$ ,  $g \equiv 0$  in the notation of that reference). We apply this to the increment  $\delta_y w$ , cf. (17), to the effect of

$$\begin{aligned} & \ell^{\alpha_0} [\nabla \delta_y w]_{\alpha_0, P_{2\ell}} + \|\nabla \delta_y w\|_{P_{2\ell}} \\ & \lesssim \ell^{-1} \inf_k \|\delta_y w - k\|_{P_{3\ell}} + \ell^{\alpha_0} [a_y \nabla \delta_y v + \delta_y j]_{\alpha_0, P_{3\ell}}. \end{aligned} \quad (43)$$

We first argue that we may upgrade (43) to

$$\begin{aligned} & \inf_c \|\nabla w - c\|_{P_r} \\ & \lesssim \sup_{|y| \leq r} \left( \ell^{-1} \inf_k \|\delta_y w - k\|_{P_{3\ell}} + \ell^{\alpha_0} [a_y \nabla \delta_y v + \delta_y j]_{\alpha_0, P_{3\ell}} \right). \end{aligned} \quad (44)$$

The first ingredient in passing from (43) to (44) is the following elementary interpolation estimate

$$\inf_c \|\nabla w - c\|_{P_r} \lesssim \sup_{|y| \leq r} (r \|\partial_t(\delta_y w)_r\|_{P_r} + \|\nabla \delta_y w\|_{P_r}), \quad (45)$$

where  $(\cdot)_r$  denotes convolution on scale  $r$  in the spatial variable. Here comes the argument for (45) where without loss of generality we may assume  $r = 1$  and restrict to estimating the first component  $\partial_1 w$  of the gradient. Given  $(t, x) \in P_1$  this follows from combining the following immediate consequences of the mean-value theorem

$$\begin{aligned} |\partial_1 w(t, x) - \partial_1 w(t, 0)| &\leq \|\nabla \delta_x w\|_{P_1}, \\ |\partial_1 w(t, 0) - (\delta_{e_1} w)_1(t, 0)| &\lesssim \sup_{|s| \leq 1} \|\nabla \delta_{se_1} w\|_{P_1}, \\ |(\delta_{e_1} w)_1(t, 0) - (\delta_{e_1} w)_1(0, 0)| &\leq \|\partial_t (\delta_{e_1} w)_1\|_{P_1}, \end{aligned}$$

so that  $c$  in (45) is given by  $(\delta_{e_1} w)_1(0, 0)$ . The second ingredient in passing from (43) to (44) is

$$\begin{aligned} r \|\partial_t (\delta_y w)_r\|_{P_r} \\ \lesssim \ell^{\alpha_0} ([\nabla \delta_y w]_{\alpha_0, P_{2\ell}} + [a \nabla \delta_y v + \delta_y j]_{\alpha_0, P_{2\ell}}) + \|\nabla \delta_y w\|_{P_{2\ell}}. \end{aligned} \quad (46)$$

In order to see this we apply the spatial convolution operator  $(\cdot)_r$  to (17) to the effect of

$$\partial_t (\delta_y w)_r = \nabla \cdot (a_y \nabla \delta_y w + a_y \nabla \delta_y v + \delta_y j)_r.$$

From this representation and  $r \leq \ell$  we obtain the estimate

$$\begin{aligned} \|\partial_t (\delta_y w)_r\|_{P_\ell} \\ &\lesssim r^{\alpha_0-1} [a_y \nabla \delta_y w + a_y \nabla \delta_y v + \delta_y j]_{\alpha_0, P_{2\ell}} \\ &\leq r^{\alpha_0-1} ([a_y]_{\alpha_0, P_{2\ell}} \|\nabla \delta_y w\|_{P_{2\ell}} + \|a_y\|_{P_{2\ell}} [\nabla \delta_y w]_{\alpha_0, P_{2\ell}} \\ &\quad + [a_y \nabla \delta_y v + \delta_y j]_{\alpha_0, P_{2\ell}}) \\ &\stackrel{(42), (14)}{\lesssim} r^{-1} \left(\frac{r}{\ell}\right)^{\alpha_0} \|\nabla \delta_y w\|_{P_{2\ell}} + r^{\alpha_0-1} ([\nabla \delta_y w]_{\alpha_0, P_{2\ell}} + [a_y \nabla \delta_y v + \delta_y j]_{\alpha_0, P_{2\ell}}), \end{aligned}$$

which yields (46) because of  $r \leq \ell$ . Inserting (43) into (46), and the outcome into (45), we obtain (44).

We now address the right-hand-side terms of (44). In view of (34) (slightly modified) we have for the first right-hand-side term

$$\inf_k \|\delta_y w - k\|_{P_{3\ell}} \leq r \sup_c \|\nabla w - c\|_{P_{4\ell}}. \quad (47)$$

We now turn to the second right-hand-side term of (44) and note that

$$\begin{aligned} [a \nabla \delta_y v + \delta_y j]_{\alpha_0, P_{3\ell}} &\leq [a_y]_{\alpha_0, P_{3\ell}} \|\nabla \delta_y v\|_{P_{3\ell}} + \|a_y\|_{P_{3\ell}} [\nabla \delta_y v]_{\alpha_0, P_{3\ell}} + [\delta_y j]_{\alpha_0, P_{3\ell}} \\ &\stackrel{(42), (14)}{\lesssim} \ell^{-\alpha_0} \|\nabla \delta_y v\|_{P_{3\ell}} + [\nabla \delta_y v]_{\alpha_0, P_{3\ell}} + [\delta_y j]_{\alpha_0, P_{3\ell}}. \end{aligned} \quad (48)$$

While obviously

$$\|\nabla \delta_y v\|_{P_{3\ell}} \leq r^\alpha [\nabla v]_{\alpha, P_{4\ell}}, \quad (49)$$

we need a little argument to see

$$[\nabla \delta_y v]_{\alpha_0, P_{3\ell}} + [\delta_y j]_{\alpha_0, P_{3\ell}} \lesssim r^{\alpha-\alpha_0} ([\nabla v]_{\alpha, P_{4\ell}} + [j]_{\alpha, P_{4\ell}}). \quad (50)$$

Indeed, let us focus on  $j$ ; given two points  $z, z'$  in  $P_{3\ell}$  we write  $\delta_y j(z) - \delta_y j(z')$  in the two ways of  $(j(z + (0, y)) - j(z)) - (j(z' + (0, y)) - j(z'))$  and  $(j(z + (0, y)) - j(z' + (0, y))) - (j(z) - j(z'))$  to see that (because of  $|y| \leq r \leq \ell$ )

$$|\delta_y j(z) - \delta_y j(z')| \leq 2[j]_{\alpha, P_{4\ell}} (\min\{d(z, z'), r\})^\alpha,$$

and thus as desired

$$\begin{aligned} \frac{|\delta_y j(z) - \delta_y j(z')|}{d^{\alpha_0}(z, z')} &\leq 2[j]_{\alpha, P_{4\ell}} \min\{d^{\alpha-\alpha_0}(z, z'), r^\alpha d^{-\alpha_0}(z, z')\} \\ &\leq 2[j]_{\alpha, P_{4\ell}} r^{\alpha-\alpha_0} \quad \text{since } \alpha \geq \alpha_0 \geq 0. \end{aligned}$$

Inserting (49) and (50) into (48) we obtain

$$\begin{aligned} [a \nabla \delta_y v + \delta_y j]_{\alpha_0, P_{3\ell}} \\ \lesssim \ell^{-\alpha_0} r^\alpha [\nabla v]_{\alpha, P_{4\ell}} + r^{\alpha-\alpha_0} ([\nabla v]_{\alpha, P_{4\ell}} + [j]_{\alpha, P_{4\ell}}). \end{aligned} \quad (51)$$

Inserting (47) and (51) into (44) we obtain the iterable form

$$\begin{aligned} \inf_c \|\nabla w - c\|_{P_r} \\ \lesssim \frac{r}{\ell} \inf_c \|\nabla w - c\|_{P_{4\ell}} + r^\alpha \left(\frac{\ell}{r}\right)^{\alpha_0} ([\nabla v]_{\alpha, P_{4\ell}} + [j]_{\alpha, P_{4\ell}}). \end{aligned}$$

Relabelling  $4\ell$  by  $\ell$  we obtain for all  $r \leq \ell \leq L$

$$\begin{aligned} r^{-\alpha} \inf_c \|\nabla w - c\|_{P_r} \\ \lesssim \left(\frac{r}{\ell}\right)^{1-\alpha} \ell^{-\alpha} \inf_c \|\nabla w - c\|_{P_\ell} + \left(\frac{\ell}{r}\right)^{\alpha_0} ([\nabla v]_{\alpha, P_L} + [j]_{\alpha, P_L}). \end{aligned}$$

By the triangle inequality in  $\|\cdot\|$  and by  $\sup_{r \leq L} r^{-\alpha} \inf_c \|\nabla v - c\|_{P_r} \leq [\nabla v]_{\alpha, P_L}$  this may be upgraded to

$$\begin{aligned} r^{-\alpha} \inf_c \|\nabla u - c\|_{P_r} \\ \lesssim \left(\frac{r}{\ell}\right)^{1-\alpha} \ell^{-\alpha} \inf_c \|\nabla u - c\|_{P_\ell} + \left(\frac{\ell}{r}\right)^{\alpha_0} ([\nabla v]_{\alpha, P_L} + [j]_{\alpha, P_L}). \end{aligned}$$

Slaving  $\ell$  to  $r$  via  $\ell = Mr$  for some  $M \geq 1$  to be chosen later, we obtain from distinguishing the ranges  $r \leq \frac{L}{M}$  and  $\frac{L}{M} \leq r \leq L$  that

$$\begin{aligned} \sup_{r \leq L} r^{-\alpha} \inf_c \|\nabla u - c\|_{P_r} &\lesssim \sup_{\frac{L}{M} \leq r \leq L} r^{-\alpha} \inf_c \|\nabla u - c\|_{P_r} \\ &+ M^{\alpha-1} \sup_{\ell \leq L} \ell^{-\alpha} \inf_c \|\nabla u - c\|_{P_\ell} + M^{\alpha_0} ([\nabla v]_{\alpha, P_L} + [j]_{\alpha, P_L}). \end{aligned} \quad (52)$$

Clearly, the first right-hand-side term is controlled as follows

$$\sup_{\frac{L}{M} \leq r \leq L} r^{-\alpha} \inf_c \|\nabla u - c\|_{P_r} \leq \left(\frac{M}{L}\right)^{\alpha-\alpha_0} [\nabla u]_{\alpha_0, P_L} \stackrel{(19)}{\leq} M^{\alpha-\alpha_0} L^{-\alpha}.$$

Hence fixing an  $M = M(d, \lambda, \Lambda, \alpha_0, \alpha)$  sufficiently large, we may absorb the second right-hand-side term in (52) into the left hand side to obtain

$$\sup_{r \leq L} r^{-\alpha} \inf_c \|\nabla u - c\|_{P_r} \lesssim L^{-\alpha} + [\nabla v]_{\alpha, P_L} + [j]_{\alpha, P_L}. \quad (53)$$

For this, we do not need to know beforehand that the left hand side side is finite, since (52) also holds when the two suprema are restricted to  $\epsilon \leq r \leq L$  and  $\epsilon \leq \ell \leq L$  for any  $\epsilon > 0$ , which is finite since  $\nabla u$  is in particular assumed to be continuous. Hence we obtain (53) with supremum restricted to  $\epsilon \leq r \leq L$ , in which we now may let  $\epsilon \downarrow 0$  to recover the form as stated in (53). By the standard norm equivalence

$$\sup_{r \leq L} r^{-\alpha} \|\nabla u - \nabla u(0)\|_{P_r} \lesssim \sup_{r \leq L} r^{-\alpha} \inf_c \|\nabla u - c\|_{P_r},$$

and shifting the origin into an arbitrary  $z \in P_L$ , we obtain (20) from (53).

## 6 Proof of Corollary 1

Throughout the proof, we use  $\lesssim$  as in Lemma 2.

By Lemma 1, the hypothesis (19) of Lemma 2 is satisfied provided we fix  $L = c([\nabla v]_{\alpha_0} + [j]_{\alpha_0})^{-\alpha_0}$  for  $c = c(d, \lambda, \alpha)$  sufficiently small. Hence we obtain from (20) that

$$[\nabla u]_{\alpha, P_L} \lesssim ([\nabla v]_{\alpha_0} + [j]_{\alpha_0})^{\frac{\alpha}{\alpha_0}} + ([\nabla v]_{\alpha} + [j]_{\alpha}).$$

By translation invariance of our deterministic setting, this persists with  $P_L$  replaced by the shifted parabolic cylinder  $P_L(z) = z + P_L$  for any point  $z \in \mathbb{R} \times \mathbb{R}^d$ , leading to

$$[\nabla u]_{\alpha, P_L(z)} \lesssim ([\nabla v]_{\alpha_0} + [j]_{\alpha_0})^{\frac{\alpha}{\alpha_0}} + ([\nabla v]_{\alpha} + [j]_{\alpha}).$$



This yields the desired Hölder estimate on  $\nabla u$  for points  $z, z'$  at parabolic distance less than  $L$ . For those  $z, z'$  with  $d(z, z') \geq L$  we appeal once more to (18) in form of

$$\begin{aligned} |\nabla u(z) - \nabla u(z')| &\lesssim ([\nabla v]_{\alpha_0} + [j]_{\alpha_0}) d^{\alpha_0}(z, z') \\ &\leq ([\nabla v]_{\alpha_0} + [j]_{\alpha_0}) L^{\alpha_0 - \alpha} d^{\alpha}(z, z') \sim ([\nabla v]_{\alpha_0} + [j]_{\alpha_0})^{\frac{\alpha}{\alpha_0}} d^{\alpha}(z, z'), \end{aligned}$$

where we used the definition of  $L$  in the last step.

It remains to estimate the  $C^{1-\alpha}$ -norm of  $w := u - v$ , more precisely, it just remains to estimate the temporal continuity, cf. (10):

$$|w(t, x) - w(t', x)| \lesssim ([\nabla u]_{\alpha} + [j]_{\alpha} + [\nabla w]_{\alpha}) \sqrt{|t - t'|}^{1+\alpha}. \quad (54)$$

To this purpose, we rewrite (8) as  $\partial_t w = \nabla \cdot (A(\nabla u) + j)$  to which we apply spatial convolution on scale  $r$  to be fixed later. This yields the estimate

$$\|\partial_t w_r\| \lesssim \frac{1}{r^{1-\alpha}} [A(\nabla u) + j]_{\alpha} \stackrel{(16)}{\lesssim} \frac{1}{r^{1-\alpha}} ([\nabla u]_{\alpha} + [j]_{\alpha}).$$

From this we deduce

$$|w_r(t, x) - w_r(t', x)| \lesssim ([\nabla u]_{\alpha} + [j]_{\alpha}) \frac{|t - t'|}{r^{1-\alpha}}.$$

We may take the convolution kernel  $\phi_r$  to be symmetric, so that in particular  $w_r(t, x) = \int \phi_r(x - y)(w(t, y) - \nabla w(t, x) \cdot (y - x)) dy$ , to the effect of

$$|w(t, x) - w_r(t, x)| \lesssim [\nabla w]_{\alpha} r^{1+\alpha}.$$

The last two estimates combine to

$$|w(t, x) - w(t', x)| \lesssim ([\nabla u]_{\alpha} + [j]_{\alpha}) \frac{|t - t'|}{r^{1-\alpha}} + [\nabla w]_{\alpha} r^{1+\alpha}.$$

Optimizing through the choice of  $r = \sqrt{t}$  yields (54).

## 7 Proof of Lemma 3

Throughout this proof we use  $\lesssim$  for  $\leq C(\alpha, s, d)$ .

Throughout the proof we fix  $j \in \{1, \dots, d\}$  and set  $h = \partial_j v$ . We aim to show that for  $C$  large enough and  $\alpha < \min\{\frac{s-d}{2}, 1\}$

$$\left\langle \exp\left(\frac{1}{C}[h]_{\alpha}\right) \right\rangle < \infty.$$

We assume without loss of generality that  $\frac{s-d}{2} < 1$ .

First we recall that by definition  $v$  and  $h$  are 1-periodic in each spatial direction and  $v(t, x) = h(t, x) = 0$  for  $t \leq 0$ . Furthermore for  $t > 1$ ,  $h$  solves

$$(\partial_t - \Delta)h = 0,$$

so that by standard continuity properties of the heat equation in Hölder norms we have  $[h]_\alpha \lesssim [h]'_\alpha$  where  $[h]'_\alpha$  is the local Hölder norm defined by

$$[h]'_\alpha := \sup_{R \in (0, 1)} \frac{1}{R^\alpha} \sup_{\substack{(t, x), (s, y) \in (0, 1) \times (-1, 1)^d \\ \sqrt{|t-s|} + |x-y| < R}} |h(t, x) - h(s, y)|.$$

We thus aim to establish

$$\left\langle \exp \left( \frac{1}{C} [h]_\alpha'^2 \right) \right\rangle < \infty. \quad (55)$$

The core stochastic ingredient for the proof of (55) is the following bound on second moments of increments of  $h$ : For  $(t, x), (t', x') \in [0, 1] \times \mathbb{R}^d$  we have

$$\langle (h(t, x) - h(t', x'))^2 \rangle \lesssim |t - t'|^{\frac{s-d}{2}} + |x - x'|^{s-d}. \quad (56)$$

The argument for (56) is based on the following Fourier representation for  $h$ : For  $t \in [0, 1]$  and  $x \in \mathbb{R}^d$  we get by differentiating (22) with respect to  $x_j$

$$h(t, x) = \sum_{k \in (2\pi\mathbb{Z})^d} \sqrt{\hat{K}(k)} i k_j e^{ik \cdot x} \int_0^t e^{-(t-s)|k|^2} d\beta_k(s),$$

which for  $t' \leq t$  leads to

$$\begin{aligned} \langle h(t, x) h(t', x') \rangle &= \sum_{k \in (2\pi\mathbb{Z})^d} \hat{K}(k) k_j^2 e^{ik \cdot (x-x')} \int_0^{t'} e^{-(t-s)|k|^2} e^{-(t'-s)|k|^2} ds \\ &= \sum_{k \in (2\pi\mathbb{Z})^d} \hat{K}(k) \frac{k_j^2}{2|k|^2} e^{ik \cdot (x-x')} [e^{-(t-t')|k|^2} - e^{-(t+t')|k|^2}]. \end{aligned} \quad (57)$$

In order to deduce (56), we use the triangle inequality and treat the cases  $t = t', x \neq x'$  and  $t \neq t', x = x'$  separately. In the first case we get using stationarity in  $x$  in the first and the symmetry of  $\hat{K}$  in the last equality

$$\begin{aligned} &\langle (h(t, x) - h(t, x'))^2 \rangle \\ &= 2 \langle h(t, x)^2 - h(t, x) h(t, x') \rangle \\ &= 2 \sum_{k \in (2\pi\mathbb{Z})^d} \hat{K}(k) \frac{k_j^2}{2|k|^2} (1 - e^{ik \cdot (x-x')}) [1 - e^{-2t|k|^2}] \end{aligned}$$

$$= \sum_{k \in (2\pi\mathbb{Z})^d} \hat{K}(k) \frac{k_j^2}{2|k|^2} (2 - e^{ik \cdot (x-x')} - e^{-ik \cdot (x-x')}) [1 - e^{-2t|k|^2}].$$

Now using the simple estimates  $\frac{k_j^2}{2|k|^2} \leq \frac{1}{2}$ ,  $|2 - e^{ik \cdot (x-x')} - e^{-ik \cdot (x-x')}| \leq \min\{4, |k|^2 |x - x'|^2\}$  as well as  $[1 - e^{-2t|k|^2}] \leq 1$ , and recalling condition (23) on  $\hat{K}$  this turns into the estimate

$$\begin{aligned} & \langle (h(t, x) - h(t, x'))^2 \rangle \\ & \lesssim \sum_{|k| \leq |x-x'|^{-1}} \hat{K}(k) |k| |x - x'| + \sum_{|k| > |x-x'|^{-1}} \hat{K}(k) \\ & \lesssim |x - x'|^2 \sum_{|k| \leq |x-x'|^{-1}} \frac{|k|^2}{(1 + |k|^2)^{\frac{s}{2}}} + \sum_{|k| > |x-x'|^{-1}} \frac{1}{(1 + |k|^2)^{\frac{s}{2}}} \\ & \lesssim |x - x'|^{s-d}, \end{aligned}$$

where we have used our assumption that  $s - d < 2$ . In the same way we get by specialising (57) to  $x = x'$  and treating the case  $t \geq t'$

$$\begin{aligned} & \langle (h(t, x) - h(t', x))^2 \rangle \\ & = \langle h(t, x)^2 + h(t', x)^2 - 2h(t, x)h(t', x)^2 \rangle \\ & = 2 \sum_{k \in (2\pi\mathbb{Z})^d} \hat{K}(k) \frac{k_j^2}{2|k|^2} [2 - e^{-2t|k|^2} - e^{-2t'|k|^2} - 2e^{-(t-t')|k|^2} + 2e^{-(t+t')|k|^2}]. \end{aligned}$$

Now using again  $\frac{k_j^2}{2|k|^2} \leq \frac{1}{2}$  as well as

$$|2 - e^{-2t|k|^2} - e^{-2t'|k|^2} - 2e^{-(t-t')|k|^2} + 2e^{-(t+t')|k|^2}| \leq 4 \min\{1, |t - t'| |k|^2\},$$

and using (23) once more this turns into

$$\begin{aligned} & \langle (h(t, x) - h(t', x))^2 \rangle \\ & \lesssim |t - t'| \sum_{|k|^2 \leq |t-t'|^{-1}} \frac{|k|^2}{(1 + |k|^2)^{\frac{s}{2}}} + \sum_{|k|^2 > |t-t'|^{-1}} \frac{1}{(1 + |k|^2)^{\frac{s}{2}}} \\ & \lesssim |t - t'|^{\frac{s-d}{2}}, \end{aligned}$$

and thus (56) follows.

We now apply Kolmogorov's continuity theorem to  $h$ ; for the convenience of the reader we give a self-contained argument. We first appeal to Gaussianity to post-process (56), which we rewrite as

$$\left\langle \frac{1}{R^{s-d}} (h(t, x) - h(s, y))^2 \right\rangle \lesssim 1 \quad \text{provided } |t - s| \leq 3R^2, |x - y| \leq R$$

for a given scale  $R$ . By Gaussianity of  $h$  we can upgrade this estimate to

$$\left\langle \exp \left( \frac{1}{C R^{s-d}} (h(t, x) - h(s, y))^2 \right) \right\rangle \lesssim 1$$

for  $|t - s| \leq 3R^2, |x - y| \leq R$ . (58)

Thus proving the desired estimate (55) on Gaussian moments of the local Hölder-norm  $[h]_\alpha'$  amounts to exchanging the expectation and the supremum over  $(t, x), (s, y)$  in (58) at the prize of a decreased Hölder exponent  $\alpha < \frac{s-d}{2}$ . To this purpose, we now argue that for  $\alpha > 0$ , the supremum over a continuum can be replaced by the supremum over a discrete set: For  $R < 1$  we define the grid

$$\Gamma_R = [0, 1] \times [-1, 1]^d \cap (R^2 \mathbb{Z} \times R \mathbb{Z}^d)$$

and claim that

$$[h]_\alpha' \lesssim \sup_R \frac{1}{R^\alpha} \sup_{\substack{(t,x), (s,y) \in \Gamma_R \\ |t-s| \leq 3R^2, |x-y| \leq R}} |h(t, x) - h(s, y)| =: \Theta,$$

where the first supremum runs over all  $R$  of the form  $2^{-N}$  for an integer  $N \geq 1$ . Hence we have to show for arbitrary  $(t, x), (s, y) \in (-1, 0) \times (-1, 1)^d$  that

$$|h(t, x) - h(s, y)| \lesssim \Theta (\sqrt{|t - s|} + |x - y|)^\alpha. \quad (59)$$

By density, we may assume that  $(t, x), (s, y) \in r^2 \mathbb{Z} \times r \mathbb{Z}^d$  for some dyadic  $r = 2^{-N} < 1$  (this density argument requires the qualitative a priori information of the continuity of  $h$ , which can be circumvented by approximating  $h$ ). For every dyadic level  $n = N, N - 1, \dots$  we now recursively construct two sequences  $(t_n, x_n), (s_n, y_n)$  of space–time points, starting from  $(t_N, x_N) = (t, x)$  and  $(s_N, y_N) = (s, y)$ , with the following properties

- a) they are in the corresponding lattice of scale  $2^{-n}$ , i. e. we have  $(t_n, x_n), (s_n, y_n) \in (2^{-n})^2 \mathbb{Z} \times 2^{-n} \mathbb{Z}^d$ ,
- b) they are close to their predecessors in the sense of  $|t_n - t_{n+1}|, |s_n - s_{n+1}| \leq 3(2^{-(n+1)})^2$  and  $|x_{n,i} - x_{n+1,i}|, |y_{n,i} - y_{n+1,i}| \leq 2^{-(n+1)}$ , where  $x_{n,i}, x_{n+1,i}, \dots$  denote the  $i$ -component of  $x_n, x_{n+1}, \dots$ . So by definition of  $\Theta$  we have

$$\begin{aligned} |h(t_n, x_n) - h(t_{n+1}, x_{n+1})| &\lesssim \Theta (2^{-(n+1)})^\alpha, \\ |h(s_n, y_n) - h(s_{n+1}, y_{n+1})| &\lesssim \Theta (2^{-(n+1)})^\alpha, \end{aligned} \quad (60)$$

and

- c) such that  $|t_n - s_n|$  and  $|x_n - y_n|$  are minimized among the points satisfying a) and b).

Because of the latter, we have

$$(t_M, x_M) = (s_M, y_M) \quad \text{for some } M \text{ with } 2^{-M} \leq \max\{\sqrt{|t-s|}, |x-y|\},$$

so that by the triangle inequality we gather from (60)

$$|h(t, x) - h(s, y)| \lesssim \sum_{n=N-1}^M \Theta(2^{-(n+1)})^\alpha \leq \Theta \frac{(2^{-M})^\alpha}{2^\alpha - 1},$$

which yields (59).

Equipped with (59), we now may upgrade (58)–(55). Indeed, (59) can be reformulated on the level of indicator functions  $I$  as

$$I\left(\left([h]_\alpha'\right)^2 \geq M\right) \leq \sup_R \max_{(t,x),(s,y) \in \Gamma_R} I\left(\frac{1}{R^{s-d}}(h(t, x) - h(s, y))^2 \geq \frac{M}{C R^{s-d-2\alpha}}\right),$$

where as in (59)  $R$  runs over all  $2^{-N}$  for integers  $N \geq 1$ . Replacing the suprema by sums in order to take the expectation, we obtain

$$\begin{aligned} & \langle I\left(\left([h]_\alpha'\right)^2 \geq M\right) \rangle \\ & \leq \sum_R \sum_{(t,x),(s,y)} \left\langle I\left(\frac{1}{R^{s-d}}(h(t, x) - h(s, y))^2 \geq \frac{M}{C R^{s-d-2\alpha}}\right) \right\rangle. \end{aligned}$$

We now appeal to Chebyshev's inequality in order to make use of (58):

$$\begin{aligned} & \langle I\left(\left([h]_\alpha'\right)^2 \geq M\right) \rangle \\ & \lesssim \sum_R \sum_{(t,x),(s,y)} \exp\left(-\frac{M}{C R^{s-d-2\alpha}}\right) \\ & \lesssim \sum_R \frac{1}{R^{2+d}} \exp\left(-\frac{M}{C R^{s-d-2\alpha}}\right) \\ & \stackrel{R \leq 1, M \geq 1}{\leq} \exp\left(-\frac{M}{C}\right) \sum_R \frac{1}{R^{2+d}} \exp\left(-\frac{1}{C} \left(\frac{1}{R^{s-d-2\alpha}} - 1\right)\right) \lesssim \exp\left(-\frac{M}{C}\right), \end{aligned}$$

where in the second step we have used that the number of pairs  $(t, x), (s, y)$  of neighboring lattice points is bounded by  $C \frac{1}{R^{2+d}}$  and in the last step we have used that stretched exponential decay (recall  $s - d - 2\alpha > 0$ ) beats polynomial growth. The last estimate immediately yields (55).

**Acknowledgements** Funding was provided by Royal Society (University Research Fellowship UF14018).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Da Prato, G., Zabczyk, J.: Stochastic Equations in Infinite Dimensions. Cambridge university press, Cambridge (2014)
2. Debussche, A., De Moor, S., Hofmanová, M.: A regularity result for quasilinear stochastic partial differential equations of parabolic type. *SIAM J. Math. Anal.* **47**(2), 1590–1614 (2015)
3. Hairer, M.: An introduction to stochastic PDES (2009). arXiv preprint [arXiv:0907.4178](https://arxiv.org/abs/0907.4178)
4. Krylov, N.V.: Lectures on Elliptic and Parabolic Equations in Hölder Spaces, Graduate Studies in Mathematics, vol. 12. American Mathematical Society, Providence (1996)
5. Krylov, N.V., Rozovskii, B.L.: Stochastic evolution equations. *J. Math. Sci.* **16**(4), 1233–1277 (1981)
6. Ladyzhenskaia, O.A., Solonnikov, V.A., Ural'tseva, N.N.: Linear and Quasi-linear Equations of Parabolic Type, vol. 23. American Mathematical Society, Providence (1988)
7. Lieberman, G.M.: Second Order Parabolic Differential Equations. World Scientific Publishing Co. Inc., River Edge (1996)
8. Otto, F., Weber, H.: Hölder regularity for a non-linear parabolic equation driven by space–time white noise. *ArXiv e-prints* (2015)
9. Pardoux, É.: Equations aux dérivées partielles stochastiques non lineaires monotones: Etude de solutions fortes de type Ito. Ph.D. thesis (1975)
10. Prévôt, C., Röckner, M.: A Concise Course on Stochastic Partial Differential Equations, vol. 1905. Springer, Berlin (2007)
11. Walsh, J.B.: An introduction to stochastic partial differential equations. In: Hennequin, P.L. (ed.) *École d'Été de Probabilités de Saint Flour XIV-1984*, pp. 265–439. Springer, Berlin (1986)